Version A

Universitat Pompeu Fabra

Networks, Crowds and Markets - Midterm Exam

Time limit: 1 hour 15 minutes

Total: 100 points

Name: ______________________________

Student Number (ID): ______________________________

Instructions

- Fill out your name and student number on this page.
- Carefully write your answers in the provided space. Use the **last two pages** for your own notes (not graded).
- Write your answers clearly. Provide key steps and reasoning; lengthy calculations are not required.
- This exam has one short quiz part and three problems.
- No electronic devices, notes, or books are allowed. Basic calculators are permitted but will not be useful here.

Part	Description	Points
Part I	Short quiz	25
Problem 1	Centralities on an undirected graph	25
Problem 2	$\mathrm{ER}(3, p)$: connectivity indicator	25
Problem 3	PageRank	25
	Total	100

Part I - Short Quiz (25 points)

Answer without long calculations. Circle or mark the correct option(s) as specified. Each sub-question is worth the indicated points.

(a) In an undirected simple graph with N vertices and L edges, the average degree equals ________________.

Choose one: (i) L/N, (ii) $2L/N$, (iii) $(N-1)p$, (iv) $L/(N-1)$ [3 pts]

Answer: $2L/N$ (ii).

(b) In $\mathrm{ER}(N, p)$ with $p = \lambda/N$ and fixed $\lambda > 0$, the degree of a fixed node is approximately ____________________ for large N.

Choose one: (i) Normal, (ii) Binomial$(N-1, p)$, (iii) Poisson(λ), (iv) Geometric [3 pts]

Answer: Poisson(λ) (iii). Here I also accepted (ii) and (ii)+(iii).

(c) In an undirected connected graph, the random walk with transition $P = D^{-1}A$ has stationary distribution proportional to ______________________.

Choose one: (i) 1, (ii) degrees, (iii) eigenvector centrality, (iv) closeness [3 pts]

Answer: degrees (ii).

(d) True/False: In $\mathrm{ER}(N, p)$, the connectivity threshold occurs around $p \approx \frac{\log N}{N}$. [4 pts]

Answer: True.

(e) Clustering in $\mathrm{ER}(N, p)$: $\mathbb{E}[C_v]$ equals __________. Choose one: (i) p, (ii) p^2, (iii) $1/N$, (iv) 0 [4 pts]

Answer: p (i).

(f) Betweenness centrality captures ______________________________.

Choose one: (i) average distance of a node, (ii) fraction of shortest paths going through a node, (iii) degree, (iv) triangles incident to a node [4 pts]

Answer: fraction of shortest paths going through a node (ii).

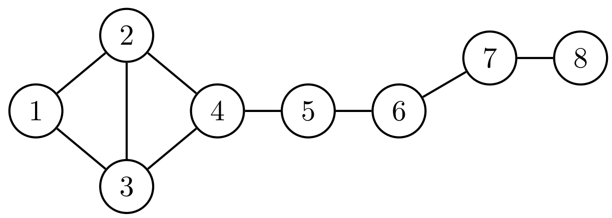
(g) As $\alpha \to 0$ in PageRank $P_\alpha = \alpha P + (1-\alpha)\frac{1}{N}\mathbf{1}\mathbf{1}^\top$, the PageRank vector tends to ________________.

Choose one: (i) degree-proportional, (ii) uniform, (iii) eigenvector centrality, (iv) undefined [4 pts]

Answer: uniform (ii).

Problem 1 - Centralities on a small undirected graph (25 points)

Consider the following simple undirected graph G on 8 nodes:



(a) Compute the degree of each node. [5 pts]
$\deg(1) = 2,\ \deg(2) = 3,\ \deg(3) = 3,\ \deg(4) = 3,\ \deg(5) = 2,\ \deg(6) = 2,\ \deg(7) = 2,\ \deg(8) = 1.$

(b) Compute the closeness centrality of nodes 4 and 6. [7 pts]
Using $C(v) = \frac{N-1}{\sum_{u \neq v} d(v,u)}$ with $N = 8$:
For node 4: distances to others are $2, 1, 1, 1, 2, 3, 4$ (sum $= 14$), so $C(4) = \frac{7}{14} = \frac{1}{2}$.
For node 6: distances $4, 3, 3, 2, 1, 1, 2$ (sum $= 16$), so $C(6) = \frac{7}{16}$ (smaller).
Here I would accept if you only computed average distances to other vertices: $\frac{1}{2}$, $\frac{16}{7}$

(c) Compute the betweenness centrality (unnormalized) of node 4. Show your path counting. [8 pts]
Node 4 is a cut-vertex between $\{1, 2, 3\}$ and $\{5, 6, 7, 8\}$. All shortest paths between any $s \in \{1, 2, 3\}$ and $t \in \{5, 6, 7, 8\}$ pass through 4. There are $3 \times 4 = 12$ such unordered pairs, and for each the fraction of shortest through 4 is 1. We also check that for any other pair $s, t \in \{1, 2, 3\}$ or $s, t \in \{5, 6, 7, 8\}$ the shortest paths between s and t do not contain 4 contributing zero to the total. Hence $B(4) = 12$ (unnormalized).

(d) Eigenvector centrality (qualitative): Between nodes 4 and 6, which one should have a higher eigenvector centrality? Give a one- or two-sentence justification. [5 pts]
Node 4. It is adjacent to the higher-score triangle nodes (2 and 3) and acts as a bridge to the tail, while node 6 connects only to degree-2 neighbours (5 and 7). Thus eigenvector centrality favours 4.
Here I accepted a wide range of answers that roughly referred to the 'importance of neighbours". I did not accept answers that only referred to other types of centrality like degree centrality, or betweenness centrality.

Problem 2 - $\mathrm{ER}(3,p)$: connectivity indicator (25 points)

Let $G \sim \mathrm{ER}(3,p)$. Let $D = \mathbf{1}\{G \text{ is disconnected}\}$ be the indicator of the event that G is disconnected.

(a) Compute $\mathbb{E}[D]$ as a function of p. [8 pts]

By definition, $D = 1$ if G is disconnected and $D = 0$ if G is connected. In the seminar we simply drew all the eight possible graphs and noted that the disconnected ones are precisely the empty graph ($L = 0$) and the graphs with one edge ($L = 1$). Thus, $D = 1$ if and only if $L \in \{0, 1\}$. With $L \sim \mathrm{Binomial}(3, p)$,

$$\mathbb{E}[D] = \Pr(L = 0) + \Pr(L = 1) = (1 - p)^3 + 3p(1 - p)^2 = 1 - 3p^2 + 2p^3.$$

(b) Compute $\mathrm{Var}(D)$. [6 pts]

D is Bernoulli with mean $\mu = 1 - 3p^2 + 2p^3$, hence

$$\mathrm{Var}(D) = \mu(1 - \mu) = (1 - 3p^2 + 2p^3)(3p^2 - 2p^3).$$

(c) Let L be the number of edges. Compute $\mathbb{E}[L]$ and $\mathrm{Var}(L)$. State the distribution of L. [4 pts]

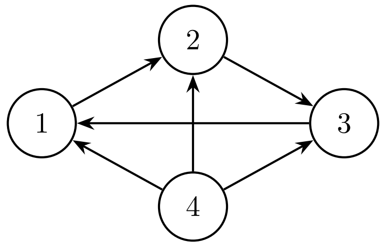
$L \sim \mathrm{Binomial}(3, p)$, so $\mathbb{E}[L] = 3p$ and $\mathrm{Var}(L) = 3p(1 - p)$.

(d) Briefly explain why, for small p, $\Pr(G \text{ is connected}) \approx 3p^2 + o(p^2)$. [7 pts]

$\Pr(\text{connected}) = \Pr(L = 2) + \Pr(L = 3) = 3p^2(1 - p) + p^3 = 3p^2 + o(p^2)$ as $p \to 0$; the two-edge (path) case dominates.

Problem 3 - PageRank (25 points)

Consider the directed network below.



The edges are:

$$1 \to \{2\}, \quad 2 \to \{3\}, \quad 3 \to \{1\}, \quad 4 \to \{1,2,3\}.$$

(a) Write the corresponding transition matrix P. [4 pts]

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

(b) Define for any $\alpha \in (0,1)$

$$P_\alpha \;=\; \alpha P \;+\; (1-\alpha)\,\frac{1}{4}\,\mathbf{1}\mathbf{1}^\top$$

and let $R = P_{0.5}$. Let $\mathbf{u} = (1,1,1,0)$. Show that $R^\top \mathbf{1} = \frac{2}{3}\mathbf{u} + \frac{1}{2}\mathbf{1}$ and $R^\top \mathbf{u} = \frac{1}{2}\mathbf{u} + \frac{3}{8}\mathbf{1}$. [8 pts]

Note that R may be a complicated matrix but both P and $\frac{1}{4}\mathbf{1}\mathbf{1}^\top$ are very simple (sparse, constant). In particular, we see immediately that

$$P^\top \mathbf{1} = \tfrac{4}{3}\mathbf{u} \qquad \text{and} \qquad P^\top \mathbf{u} = \mathbf{u}$$

and

$$(\tfrac{1}{4}\mathbf{1}\mathbf{1}^\top)\mathbf{1} = \tfrac{1}{4}\mathbf{1}(\mathbf{1}^\top\mathbf{1}) = \mathbf{1} \qquad \text{and} \qquad (\tfrac{1}{4}\mathbf{1}\mathbf{1}^\top)\mathbf{u} = \tfrac{1}{4}\mathbf{1}(\mathbf{1}^\top\mathbf{u}) = \tfrac{3}{4}\mathbf{1}.$$

Since $R^\top = \frac{1}{2}P^\top + \frac{1}{2}\frac{1}{4}\mathbf{1}\mathbf{1}^\top$, we get

$$R^\top\mathbf{1} = \tfrac{1}{2}P^\top\mathbf{1} + \tfrac{1}{2}\tfrac{1}{4}\mathbf{1}\mathbf{1}^\top\mathbf{1}\,\mathbf{1} = \tfrac{1}{2}\big(\tfrac{4}{3}\mathbf{u}\big) + \tfrac{1}{2}\mathbf{1} \;= \tfrac{2}{3}\mathbf{u} + \tfrac{1}{2}\mathbf{1}$$

and

$$R^\top\mathbf{u} = \tfrac{1}{2}P^\top\mathbf{u} + \tfrac{1}{2}\tfrac{1}{4}\mathbf{1}^\top\mathbf{1}\,\mathbf{u} = \tfrac{1}{2}\mathbf{u} + \tfrac{1}{2}\cdot\tfrac{3}{4}\mathbf{1} = \tfrac{1}{2}\mathbf{u} + \tfrac{3}{8}\mathbf{1}.$$

(c) Starting from the uniform row vector $\pi^{(0)} = \frac{1}{4}\mathbf{1}$, perform **two** power-iteration steps given for $k \geq 0$ by

$$\pi^{(k+1)} = R^\top \pi^{(k)}.$$

In other words, compute $\pi^{(1)}$ and $\pi^{(2)}$. Hint: Use (b). [8 pts]

Note that this could be done even if, for some reason, you could not solve (a) and (b). Just follow the hint and use the explicit formulas that were given to you in (b): $R^\top \mathbf{1} = \frac{2}{3}\mathbf{u} + \frac{1}{2}\mathbf{1}$ and $R^\top \mathbf{u} = \frac{1}{2}\mathbf{u} + \frac{3}{8}\mathbf{1}$.

In particular:

$$\pi^{(1)} = R^\top \pi^{(0)} = R^\top \frac{1}{4}\mathbf{1} = \frac{1}{4}R^\top \mathbf{1} = \frac{1}{4}\left(\frac{2}{3}\mathbf{u} + \frac{1}{2}\mathbf{1}\right) = \frac{1}{6}\mathbf{u} + \frac{1}{8}\mathbf{1}$$

and

$$\pi^{(2)} = R^\top \pi^{(1)} = \frac{1}{6}R^\top \mathbf{u} + \frac{1}{8}R^\top \mathbf{1} = \frac{1}{6}\left(\frac{1}{2}\mathbf{u} + \frac{3}{8}\mathbf{1}\right) + \frac{1}{8}\left(\frac{2}{3}\mathbf{u} + \frac{1}{2}\mathbf{1}\right) = \frac{1}{6}\mathbf{u} + \frac{1}{8}\mathbf{1} = \pi^{(1)}.$$

This shows that $R^\top \pi^{(1)} = \pi^{(1)}$ and so $\pi^{(1)}$ is teh stationary distribution that gives the PageRank ranking! Also note that

$$\pi^{(1)} = \begin{bmatrix} \frac{1}{6} + \frac{1}{8} \\ \frac{1}{6} + \frac{1}{8} \\ \frac{1}{6} + \frac{1}{8} \\ \frac{1}{8} \end{bmatrix} = \begin{bmatrix} \frac{7}{24} \\ \frac{7}{24} \\ \frac{7}{24} \\ \frac{3}{24} \end{bmatrix}$$

(d) Based on the calculations in (c), which nodes are the most important in PageRank? Hint: (c) gives already the stationary distribution of R. Why? [5 pts]

Nodes $1, 2, 3$ tie as most important: $\pi^* = (7/24, 7/24, 7/24, 3/24)$.

Rough Notes (not graded)

Use this page for scratch work.

Rough Notes (not graded)

Use this page for scratch work.